

# Application of artificial neural networks in the geographical identification of coffee samples

Dionísio Borsato · Marcos Vinícios Roberto Pina ·  
Kelly Roberta Spacino · Maria Brígida dos Santos Scholz ·  
Armando Androcioli Filho

Received: 4 March 2011 / Revised: 11 July 2011 / Accepted: 20 July 2011 / Published online: 2 August 2011  
© Springer-Verlag 2011

**Abstract** The recognition of the characteristics of coffee associated with a given agricultural system and aimed at adding value and attending the consumers' demands stimulates the production of types of coffee properly described. The objective of this study was to explore and to explain the physicochemical characteristics and sensory attributes of the coffee grown in Parana State (Southern Brazil) based on an integrated approach of the terrior and the application of artificial neural network. Physicochemical variables of green coffee beans and roasted coffee beans were determined, as well as sensory attributes of the beverage. One hundred and seventy-two coffee samples were analyzed for moisture, proteins, chlorogenic acids, tannins, total acidity, total lipids, caffeine, total and reducing sugars and minerals. These properties were tabulated and presented to artificial neural network multilayer perceptron to be identified as the region and the city of planting. The artificial neural network classified correctly and tested 100% of the samples grown by region. For the database containing information by city, the automatic mode of the software Statistica 9.0 was used. The neural network showed 99% accuracy in training and 100% accuracy in the stage of testing and validation.

**Keywords** Artificial neural network · Multilayer perceptron · Coffee quality · Coffee aroma

## Introduction

Pattern recognition and classification is a task naturally carried out by human beings, due to the evolution and adaptation, of our central nervous system, more specifically the cerebral cortex, over thousands of years. However, solving classification problems in most cases is extremely complex especially if the patterns are described by a large number of independent variables. When this occurs, automated systems must be used [1, 2]. There are several traditional and tested statistical methodologies used to classify and recognize patterns. One of the most traditional multivariate methods is principal component analysis (PCA). In this technique, highly dimensional data are reduced to a system of low-dimension linear coordinates in order to facilitate tendency or clustering visualization [3]. However, some disadvantages of the method are well known [4]. First, it is assumed that the data can be described by linear combinations. Consequently, non-linear systems will not be well represented. A second critical point in PCA is the quality of the result that can be influenced by discrepant samples. In addition, there is a possibility that after transformations, the number of significant components may still be high that makes it difficult to extract useful information from the data [2].

Artificial neural networks (ANN) are a set of techniques based on not very conventional statistical principles that have been gaining space in pattern recognition and classification [1]. The origin of the artificial neural network theory dates back to the mathematical and engineering models of biological neurons [5]. The neuron is the basic cell unit of the brain and is the center of the human nervous system [2, 6]. The capacity for knowledge adaptation and storage in an adult brain can be attributed to two mechanisms: creation of new synapsis connections among the

---

D. Borsato (✉) · M. V. R. Pina · K. R. Spacino  
Department of Chemistry, State University of Londrina,  
P. O. Box 6001, Londrina, PR 86051-990, Brazil  
e-mail: dborsato@uel.br

M. B. S. Scholz · A. Androcioli Filho  
Parana Agronomic Institute (IAPAR),  
P. O. Box 481, Londrina, PR 86001-970, Brazil

neurons and modification of the existing synopsis. Thus, knowledge is defined as stored information or models used by a person or machine to interpret, predict and respond appropriately to the outside world [2].

In traditional computer algorithms, knowledge is constructed explicitly in the form of a priori defined rules. The artificial neural networks are computer techniques performed from a mathematical model inspired on the structure and functioning of the natural neural networks and are systems that can make generalizations, that is, they generate their own rules, to associate the input and output variables, after learning with training data [7]. This is possible by a learning algorithm that changes the weights of the connections among the neurons in function of responses obtained from the input variables [2].

Artificial neural networks seek to model, even though primitively, the logic operations by which the brain executes the most varied tasks [8, 9]. For this, processing units are created with functions similar to those of the biological neuron. Each artificial neuron can receive several input stimuli and generate several output stimuli. This stimuli flow is propagated by a connection network, as in the biological model. The propagation intensity of the stimuli between a determined connection is made by weighting [2, 6].

For the neural networks that have different forms of representation, from the inputs to the internal parameters, the development of a satisfactory solution is a real project challenge. The convergence of the method is much related to the topography of the network and the parameters adopted in the training algorithm. Depending on the choice, learning can be fixed in a minimum location or slowly converged [7]. Furthermore, in systems that work in real time, it is important to reduce the size of the training sample by removing redundant data to increase the processing speed [10].

A pre-established set of well-defined rules for the solution of a learning problem is called a learning algorithm. As can be expected, there is no single learning algorithm for the neural network project. Instead, there is a set of tools represented by a variety of learning algorithms, each offering specific advantages. Basically, the learning algorithms differ in the form in which the fit is formulated of a synopsis weight of a neuron [2]. Because artificial neural networks can solve problems of a general nature, such as approximation, classification, categorization and prediction [11], this set of techniques has been applied to a vast range of areas, especially to process control, satellite navigation, weather forecasting, signal processing, speech recognition, waste treatment, ceramic engineering, geographic origin, fire detection, the financial market and pattern recognition [6, 9, 12–15].

The multilayer perceptron (MLP) type of artificial neural network is widely used for pattern classification [8].

The architecture of this type of network consists of an input layer with a neuron for each input variable. One or two intermediate layers, which separate the patterns by forming decision frontiers, contain a quantity of neurons to be defined, and an output layer that constructs linear combinations of decision frontiers formed by hidden neurons, where the quantity of neurons depends on how many products will be classified and how they will be represented [2]. In each, one of the neurons of the MLP network is realized a weighted sum of the synopsis weights of the signs from the neurons of the previous layer. This sum, called the induced local field, is applied to a non-linear activation function that will produce the output of the neurons [2].

ANN are extremely versatile to map the complex and non-linear relationships among multiple input and output variables. The disadvantages of the method include the need for a large quantity of training data [1], difficulty in choosing the training parameters and the type of network most suitable to the problem, and variable results due to initialization and sampling [16]. However, several studies have presented promising proposals to solve or diminish the disadvantages associated with the ANN [17–19], and new softwares are implementing ANN in an automated manner, without the need to choose the training parameters, thus pointing to a better network for the proposed problem [20].

The importance of coffee in the world economy cannot be underestimated. Business related to coffee is extremely important for world trade and loses in value only to petroleum [14]. Furthermore, it is the raw material of one of the most accepted beverages in the world [21]. A beverage is produced from roasted and ground coffee that presents much appreciated aroma and flavor [22]. An interesting characteristic of the coffee beverage is the fact that it does not have a relevant nutritional value and is consumed basically due to its physiological and psychological effects related to the presence of caffeine and especially because of the pleasure and satisfaction that its aroma and flavor are able to give [23].

Coffee quality is directly related to the various physical and chemical constituents responsible for the aroma and characteristics of the beverages. There have been several attempts to associate the composition of green coffee with beverage quality. Green coffee does not have the aroma and flavor typical of the coffee beverage, and thus, roasting is essential to produce the compounds that confer the characteristics that everyone understands as coffee [24].

Among the compounds that confer the sensorial attributes of coffee, the sugars are outstanding and vary in the coffee beans according to the stage of ripeness [24] and react with the proteins that are responsible for the formation of most of the aromatic compounds of roast coffee

[25]. Phenolic compounds present in the coffee bean range from 8 to 9% depending on the degree of ripeness of the cherries and coffee beverages. Chlorogenic acids, currently understood as compounds beneficial to health, are found in variable quantities among cultivars and are more highly concentrated in completely unripe beans [26].

Caffeine is the main alkaloid in coffee to which is attributed the activity stimulation of coffee. Unripe beans are metabolically more active in caffeine synthesis than ripe beans. The caffeine concentration ranges from 0.8 to 1.9%  $m/m$  in the *Coffea arabica* species, while in the *C. robusta* species, the caffeine concentration is higher, ranging from 1.2 to 4%  $m/m$  [24].

Generally, all ripe Arabic coffee beans still on the plants have the ideal composition to obtain an excellent beverage quality. Beans depreciate almost always during the drying process where biochemical reactions resulting from enzyme action and microorganisms can alter the initial aroma and flavor characteristics [26]. Climatic conditions confer special attributes to the beverage, body, acidity and aroma of the coffee from each region. Parana has climate and soil diversity so that the most various types of coffee can be produced. When quality can be associated with production location, the aggregated value of the coffee is increased. However, in this process, there must be mechanisms that reliably prove the geographic origin of the coffee [26].

Considering the commercial importance of coffee and the efficacy of the ANN technique, it was proposed to apply and adapt the artificial neural network methodology to identify coffees produced in different cities and regions of Parana.

## Materials and methods

### Coffee samples

Samples were collected from 16 cities distributed in the coffee region of the state of Parana, including coffee plantations from the regions with clay and sandy soil. About 3-kg coffee was collected from each location and taken to Londrina on the same day that they were collected. At the Paraná Agronomic Institute-Londrina (IAPAR), the samples were immediately placed in wooden boxes with a mesh bottom and moved eight times a day until beans reached 11–12% moisture and after the samples were benefiting from it (removal of hull and parchment). For the physicochemical analysis, coffee beans were frozen with liquid nitrogen and were ground in the mill disk (model Perten 3600) with 0.6 mm final particle size. The samples were kept frozen in a freezer and thawed for analysis.

### Coffee beans size

After beneficiation, the coffee beans size of the samples was determined in a size 16 sieve appropriate for coffee, and the result was expressed in percentage of beans retained in this sieve [27].

### Physicochemical analysis

Moisture was determined by heating in an oven with air circulation at 105 °C for 3 h. Caffeine was extracted with magnesium oxide and determined by the method espectralofotométrico [28], and total chlorogenic acids were evaluated according to the methodology proposed by Clifford and Wight [29]. Total tannins were determined with Folin–Ciocalteu reagent using gallic acid as padrão [30]. Total sugar sucrose and reducing sugars were extracted with water at 70–80 °C and determined with the Somogyi and Nelson reagent [31]. Protein, titratable acidity and total lipids were determined by the respective method proposed by AOAC [30].

### Mineral analysis

The minerals P, Ca, Mg, Cu, Zn, B, Mn, Fe and S were determined in a Thermo Jarrel optical emission spectrometer by Inductively Coupled Plasma (ICP-OES) after nitroperchloric digestion. The potassium was determined by flame photometry in a Micronal photometer. The mineral patterns were prepared by stock dilutions [32].

### Sensory analysis

The coffee was roasted at roaster (Rod Bel) 210–200 °C for 8–10 min. The end of the roasting was determined by weight loss and visual color of coffee beans. An infusion of ground coffee (0.6 mm) with water (97–99 °C) in the proportion of 70 g  $L^{-1}$  was prepared in glass cups coded with three digits to be served to tasters. The eight professional coffee cup-tasters experienced in coffee evaluation for 3 years or more were trained to differentiate and quantify the quality of the beverage in the main sensory attributes (aroma intensity of attributes, quality of aroma, acidity, bitterness, astringency and body). In training, the tasters received samples with extreme quality attributes in order to differentiate the intensity of attributes. In each session, five samples were evaluated and the intensity of attributes was marked in the unstructured scale of 10 cm, where the extremes represent the minimum and maximum intensities of each attribute.

## Artificial neural networks

The automatic and manual modules were used of the Statística<sup>®</sup> software version 9.0 [20] and a multilayer perceptron neural network (MLP) containing only one hidden neuron layer.

The values for moisture content, proteins, chlorogenic acids, total tannins, titratable acidity, total lipids, caffeine, total and reducing sugars, in percentage and phosphorus, potassium, calcium, magnesium, sulfur in  $\text{g kg}^{-1}$  sample, copper, zinc, boron, manganese, iron in  $\text{mg kg}^{-1}$  were tabled and presented to the artificial neural networks.

## Databank

The databank of analyses made available by the Parana Agronomic Institute (IAPAR) consisted of a total of 172 samples from 16 different cities from Parana State. The quantity of samples was not the same for all the cities; therefore, there were cities with insufficient samples for the training, testing and validation.

## Results and discussion

A multilayer perceptron network (MLP) was used, consisting of an input layer with one neuron for each one of the 20 input variables and a single intermediate layer that separated the patterns.

The 16 cities were grouped into three regions, A, B and C, generating two cases, classification by region and by city. The distance, mean temperature and rainfall index of the cities were considered in this classification, that is, the similar climatic characteristics among the cities (Table 1).

For the first case studied, classification by regions, the artificial neural network used was formed by eight neurons in the hidden layer, with 120 training cycles. The learning process was finalized when the synapsis weights and the mean quadratic error converged to a minimum value.

Before starting the process, all the variables were transformed in a scale from zero to 1 (minimax) with a

**Table 1** Division of the samples of the cities and regions

Cities		
Region A	Region B	Region C
Francisco Alves	Nova Esperança	Guapirama
Jesuítas	Jundiá do Sul	Joaquim Távora
São Jorge do Patrocínio	Londrina	Carlópolis
Brasilândia do Sul	Uraí	Jacarezinho
Iporã	Santa Mariana	Abatiá
		Conselheiro Mairinck

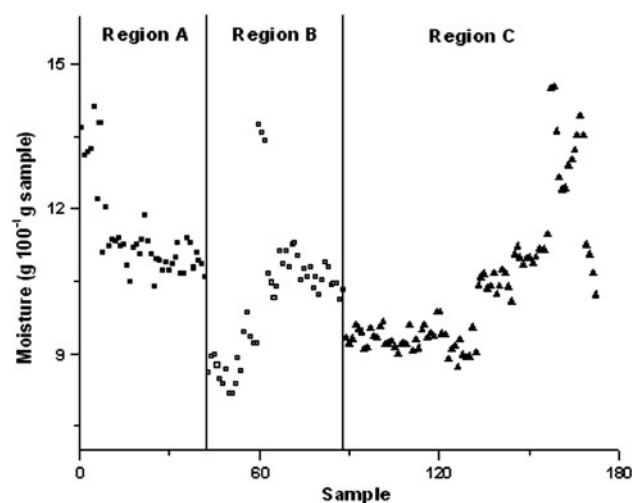
logistic-type sigmoid activation function that, according to Basheer and Hajmeer [33], is one of the most used functions in classification processes by artificial neural networks. The output layer contained three neurons, one for each region studied. The BFGS algorithm, proposed individually by Broyden–Fletcher–Goldfarb–Shanno [34–37], a quasi Newton fast convergence method and very efficient in optimization without restrictions [38], was used in the neural network constructed. The training was randomized for each cycle, and the synapsis weight was corrected based on the sum of the quadratic errors and carried out after training each sample.

The sampling was divided into three parts where one, containing 80% of the sample, was used in the network training, another with 10% was for the test stage and 10% for the validation. The samples included in this division were chosen randomly by the [20] STATISTICA 9.0<sup>®</sup> software. The validation step has the objective of verifying the capacity of the network to carry out generalizations from the trained network, since the artificial neural networks learn a rule using the training samples.

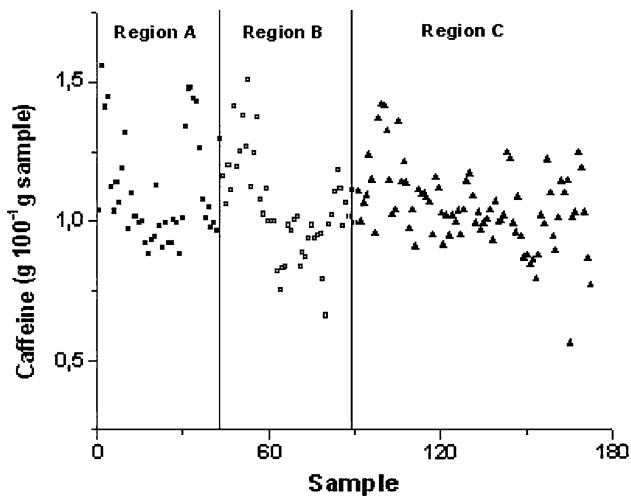
An order of importance of the variables was obtained from the trained network. The moisture content was the variable identified by the artificial neural networks as the most important in the classification, followed by the sulfur and magnesium contents, reducing sugars, lipids and proteins. Figure 1 shows the moisture content distribution by region studied.

The total sugars, zinc and caffeine contents were the variables that least influenced the classification due to the smaller variability in the data per region studied, compared to the other parameters, as can be observed in Fig. 2.

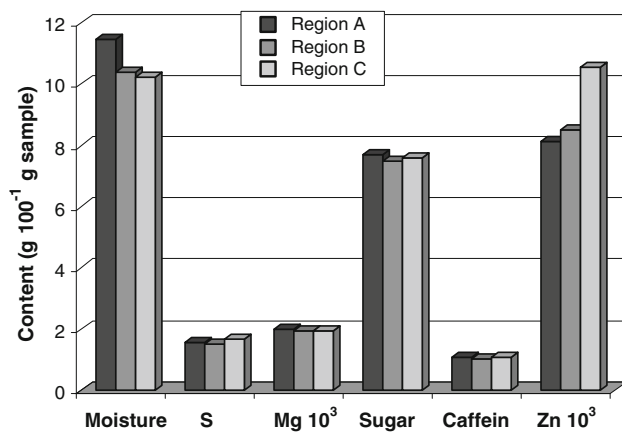
Figure 3 shows the mean moisture, sulfur and magnesium contents, variables considered the most important by



**Fig. 1** Moisture content ( $\text{g } 100^{-1}$  g sample) of the 172 coffee samples grouped in three producing regions in Parana



**Fig. 2** Caffeine content ( $\text{g } 100^{-1} \text{ g sample}$ ) of the 172 coffee samples grouped in three producing regions in Parana



**Fig. 3** Mean value of the moisture content ( $\text{g } 100^{-1} \text{ g sample}$ ), sulfur ( $\text{g } 100^{-1} \text{ g sample}$ ), magnesium ( $\text{mg } 100^{-1} \text{ g sample}$ ), total sugars ( $\text{g } 100^{-1} \text{ g sample}$ ), caffeine ( $\text{g } 100^{-1} \text{ g sample}$ ) and zinc ( $\text{mg } 100^{-1} \text{ g sample}$ ) of the samples from regions A, B and C

the network and the total sugars, caffeine and zinc contents considered the least important for classifying the samples from regions A, B and C.

In each region formed according to the conditions of climate and soil, samples of coffee had different physico-chemical and mineral compositions.

The Tukey test applied to all media of moisture content showed that there is no significant difference in level of 5%; however, the samples from regions B and C have significant differences on the same level of significance. The contents of total sugars and caffeine showed no significant difference between the means of the regions studied. Among the minerals studied, for the sulfur content was found significant difference between regions B and C at the same level of significance, whereas for the magnesium content was not found significant difference in the

three regions. For the zinc content, the Tukey test indicated that there is a significant difference only between the regions of samples A and B. However, the network did not take into consideration the analysis of the mean value of any parameter. It was trained, using a data set, to identify the characteristics and thus classify the samples [20].

The trained artificial neural networks generated an excellent response precision, presenting 99.28% accuracy in training and 100% accuracy at the test and validation stages. That is, of the 172 samples used for classification, the trained artificial neural networks only made one mistake in one classification.

For the second case, classification by city, the automatic module of the Statistica® software [20] was used in which the multilayer perceptron type of artificial neural network is used from the desired input and the output variables, in this case cities. The automatic module gets the best artificial neural network for the problem and determines the learning algorithm, activation function, number of neurons and number of cycles. For this, it was necessary to establish limits of numbers of neurons in the hidden layer, thus decreasing the convergence time of the network. The lower and upper limits of neurons in the artificial neural networks were 10 and 40, respectively, and the division of the sample followed the first case, 80% for training, 10% for the test and 10% for the validation.

The learning process of the artificial neural networks is similar to the manual module, but with *sweeping*, passing through the activation functions available in the software at an interval of the number of neurons, which results in a greater delay in obtaining the network; for more complicated cases, with a greater number of outputs, this methodology results in a good approximation to the ideal network, which can be generated later in the manual model with less convergence time since it limits the parameters.

For this case, 147 samples were used from the cities of Abatiá, Brasilândia do Sul, Francisco Alves, Guapirama, Jacarezinho, Joaquim Távora, Londrina, Nova Esperança, Santa Mariana, São Jorge do Patrocínio and Uraí. In this case, the neural network had 11 outputs because samples were taken from five cities which have insufficient number for training, testing and validation. The artificial neural network, in this case, presented 11 outputs because samples were removed from five cities with an insufficient quantity for training, testing and validation.

By the automatic module, five artificial neural networks were obtained and the best network, presenting 99% accuracy in training and 100% in the test and validation stages, was obtained with the BFGS algorithm, 18 neurons in the hidden layer, logistic function as activation function and 150 training cycles.

The input variable most important for classification in the trained artificial neural network was the quantity of

sulfur in the sample, followed by the moisture content, phosphorus and calcium quantities. The quantity of iron was the least important in the classification, followed by the boron quantity and titratable acidity.

To confirm that the artificial neural networks trained by the automatic module presented repeatability, an artificial neural network was analyzed by the manual module, choosing as parameters those obtained in the automatic module. The precision of the network generated manually was the same as that obtained in the automatic module, and the same order of importance of the input variables was also observed.

Table 2 shows the average value, the number of samples, standard deviation and significance level of the parameters that most influenced the network used in the geographical classification of the green coffees produced in the cities of Parana–Brazil.

The input parameters that most influenced the coffee aroma and flavor were the sugar, protein and chlorogenic acid contents [39]. The quality of coffee as a beverage is highly dependent on its aroma and flavor, characteristics that determine its commercial value [39]. The aroma and flavor of the coffee beverage are highly complex, resulting from the combined presence of several volatile and non-volatile chemical constituents, especially the acids, aldehydes, ketones, sugars, proteins, amino acids, fatty acids, phenolic compounds and also the enzyme actions in some of these constituents that give reaction products whose compounds interfere in the beverage flavor. Among these compounds, 29 were identified as mainly accounting for the characteristic aroma of roasted and ground coffee [40].

Complex biochemical mechanisms are involved in the production of the characteristics of coffee color, aroma and flavor during roasting such as the Maillard and Strecker reactions, sugar caramelization, chlorogenic acid degradation, protein and polysaccharide degradation [41].

Proteins, the sources of most of the aroma and flavor characteristics of coffee, are denatured when roasted at temperatures lower than pyrolysis and hydrolysis occurs in the peptide links of the protein molecules [42].

Sugars resulting from the breakage during roasting react with free amino acids and amino protein groups (Maillard reaction) to form aminoketones and aminoaldoses, which after other reactions result in numerous volatile and odorous compounds. As the roasting continues, there are other reactions involving degradation by interaction and breakage of carbohydrates and nitrogen compounds, chlorogenic acids and other acids [43]. Sucrose, fructose and glucose are the main sugars found in coffee, and their concentration is related to the maturity stage of the bean. Sucrose increases as the bean ripens, while glucose and fructose decrease at the end of the cycle [24, 44, 45]. The sugars are among the most important precursors of the aromatic

compounds formed during roasting [43] and were assessed in the sensorial analysis of the coffee beverage.

Table 3 shows the average value, the number of samples, standard deviation and significance level of the sensory attributes of coffee beans analyzed.

The sensory attributes of coffee analyzed were of little importance to classify the regions or cities by the neural network, which can be attributed to the small variability in the results (Table 3). But some physicochemical parameters interfere with their formation. By applying response surface, analysis is possible to understand the behavior of these attributes with the variation of these parameters.

Figure 4 shows the surface response relating the coffee aroma to the protein and total sugar content regardless of the producing region. In this figure, the best response for aroma was verified when the protein content presented its highest level and the total sugar content was at intermediate level (8% mass). However, the same happened when we had the upper limit for total sugars and intermediate protein levels.

The presence of phenolic compounds in coffee is responsible for astringency and interferes in the flavor. A large number of phenolic compounds have been identified in roast coffee, and some of them originate from the chlorogenic acids and account in part for the aroma and bitterness of the beverage [46, 47].

Figure 5 shows the surface response that relates the beverage aroma to the total sugars and chlorogenic acids regardless of the producing region and highlights the contribution of the chlorogenic acids to aroma formation. High coffee aroma and flavor values are found in the presence of high chlorogenic acid concentrations and high total sugar values. In the roasting condition applied, the concentration of compounds formed by the total sugar degradation may have been sufficiently high to mask the compounds resulting from the chlorogenic acid degradation responsible for the acidity and bitterness of the beverage.

Figure 6 shows that the presence of chlorogenic acid gives bitterness to the beverage in situations with low total sugar content. The presence of this sensorial attribute in the beverage is associated mainly with the degradation compounds of the chlorogenic acids formed during roasting [48, 49].

The presence of phenolic compounds represented by the chlorogenic acids exercised a strong influence on the global quality of the beverage. Although high aroma and acidity scores have been found in the presence of high values of chlorogenic acids, in the global assessment of the beverage, coffees with low chlorogenic acid contents and high total sugars values were best assessed (Fig. 7). This combination of compounds is found in ripe beans and confirms the importance of the stage of ripeness on the quality of the coffee beverage [50].

**Table 2** Average parameters of the artificial neural network used in the classification by regions and cities

Region	Cities	Moisture (%)	S (g/kg)	Proteins (%)	Mg (g/kg)	Ca (g/kg)	Lipids (%)
A	Francisco Alves ( <i>n</i> = 9)						
	Mean	12.9400 <sup>a,b</sup>	1.5322 <sup>c,d,e,f</sup>	12.6530 <sup>a,c</sup>	2.0343 <sup>a,b,c,d,e</sup>	1.3467 <sup>c,d</sup>	10.6588 <sup>f</sup>
	SD	0.9734	0.1083	1.3183	0.1206	0.0897	0.7858
	São Jorge Patrocínio ( <i>n</i> = 17)						
	Mean	11.0679 <sup>c</sup>	1.5971 <sup>c,d</sup>	12.748605 <sup>a,c</sup>	1.9741 <sup>b,c,d,e</sup>	1.4454 <sup>c</sup>	12.6418 <sup>c,d</sup>
	SD	0.3581	0.0749	0.8280	0.2269	0.1771	0.5199
	Brasílandia do Sul ( <i>n</i> = 10)						
	Mean	10.9766 <sup>c</sup>	1.5810 <sup>c,d,e</sup>	14.0221 <sup>b</sup>	2.0070 <sup>a,b,c,d,e</sup>	1.2790 <sup>c,d</sup>	11.9154 <sup>d,e,f</sup>
	SD	0.2762	0.3456	0.3503	0.4254	0.3039	0.3474
	Region A ( <i>n</i> = 42)						
	Mean	11.44507 <sup>1</sup>	1.5812 <sup>1,2</sup>	13.1274 <sup>1</sup>	2.0054 <sup>1</sup>	1.3717 <sup>2</sup>	11.9835 <sup>1</sup>
	SD	0.5637	0.1934	0.9474	0.2767	0.2110	0.7897
B	Nova Esperança ( <i>n</i> = 20)						
	Mean	9.5420 <sup>d,e,f</sup>	1.4645 <sup>c,d,e,f</sup>	12.6770 <sup>c</sup>	2.0218 <sup>a,b,c,d</sup>	1.3755 <sup>c,d</sup>	11.4505 <sup>e,f</sup>
	SD	1.7697	0.0525	0.8749	0.0960	0.1513	1.0278
	Londrina ( <i>n</i> = 20)						
	Mean	10.6665 <sup>c</sup>	1.3560 <sup>f</sup>	11.9941 <sup>c</sup>	1.9225 <sup>b,c,d,e</sup>	1.0900 <sup>d</sup>	13.9677 <sup>a</sup>
	SD	0.3267	0.0883	0.5604	0.1587	0.1686	1.0301
	Uraí ( <i>n</i> = 11)						
	Mean	10.5014 <sup>e</sup>	1.5900 <sup>c,d,e</sup>	11.7522 <sup>c,d</sup>	1.9755 <sup>b,c,d,e</sup>	1.3327 <sup>c,d</sup>	12.7970 <sup>b,c,d</sup>
	SD	0.1660	0.1485	1.4376	0.1244	0.2442	0.3958
	Santa Mariana ( <i>n</i> = 12)						
	Mean	10.9569 <sup>c</sup>	1.8117 <sup>b</sup>	10.8034 <sup>d</sup>	1.8392 <sup>c,d,e</sup>	2.1542 <sup>b</sup>	12.6910 <sup>b,c,d</sup>
	SD	0.3023	0.1368	0.8276	0.1214	0.4768	0.4271
Region B ( <i>n</i> = 69)							
Mean	10.3598 <sup>2</sup>	1.4994 <sup>2</sup>	11.9326 <sup>2</sup>	1.9285 <sup>1</sup>	1.3888 <sup>2</sup>	12.7662 <sup>2</sup>	
SD	1.1195	0.1951	1.0910	0.1572	0.4537	1.3016	
C	Guapirama ( <i>n</i> = 10)						
	Mean	9.3489 <sup>d,e,f</sup>	1.4750 <sup>c,d,e,f</sup>	12.3764 <sup>c</sup>	1.9690 <sup>b,c,d,e</sup>	1.3710 <sup>c,d</sup>	13.1417 <sup>a,b,c,d</sup>
	SD	0.1625	0.0311	0.6219	0.1222	0.2824	0.6503
	Joaquim Távora ( <i>n</i> = 9)						
	Mean	9.2752 <sup>f</sup>	1.4000 <sup>d,e,f</sup>	11.9984 <sup>c,d</sup>	1.8767 <sup>b,c,d,e</sup>	1.0011 <sup>d</sup>	13.8941 <sup>a,b</sup>
	SD	0.2092	0.0534	0.8953	0.0490	0.0874	0.7081
	Jacarezinho ( <i>n</i> = 18)						
	Mean	9.2636 <sup>f</sup>	1.4217 <sup>e,f</sup>	12.1262 <sup>c</sup>	1.8506 <sup>d,e</sup>	1.2937 <sup>c,d</sup>	12.1440 <sup>d,e</sup>
	SD	0.3165	0.0799	0.5239	0.1323	0.3203	0.9329
	Abatiá ( <i>n</i> = 11)						
	Mean	13.1133 <sup>a</sup>	2.2700 <sup>a</sup>	12.0902 <sup>c</sup>	2.0873 <sup>a,b,c</sup>	2.6600 <sup>a</sup>	13.5502 <sup>a,b,c</sup>
	SD	0.8686	0.1358	0.4075	0.1937	0.1657	0.9468
Region C ( <i>n</i> = 61)							
Mean	10.2227 <sup>2</sup>	1.6679 <sup>1</sup>	12.0811 <sup>2</sup>	1.9585 <sup>1</sup>	1.6763 <sup>1</sup>	13.1602 <sup>2</sup>	
SD	1.6729	0.4002	0.6257	0.2024	0.7835	1.0831	

For cities, means in the same column followed by same letter are not statistically different at the level of 5%

For regions, means in the same column followed by same number are not statistically different at the level of 5%

**Table 3** Main parameters used by the artificial neural network classification by regions and cities

Region	Cities	Acidity	Astringency	Bitterness	Aroma	Body	Quality aroma
A	Francisco Alves ( <i>n</i> = 9)						
	Mean	3.8038 <sup>a</sup>	4.1692 <sup>a</sup>	3.6777 <sup>a</sup>	5.6098 <sup>a</sup>	4.1061 <sup>a,b</sup>	4.9929 <sup>a</sup>
	SD	0.7123	0.4331	1.294	0.5761	0.3883	0.6099
	São Jorge do Patrocínio ( <i>n</i> = 17)						
	Mean	3.6860 <sup>a</sup>	3.9402 <sup>a,b</sup>	2.9812 <sup>a</sup>	5.3955 <sup>a</sup>	4.1268 <sup>a,b</sup>	4.9671 <sup>a</sup>
	SD	0.8048	0.4089	1.007	0.3655	0.4187	0.5323
	Brasílandia do Sul ( <i>n</i> = 10)						
	Mean	3.9285 <sup>a</sup>	3.5154 <sup>a,b,c</sup>	2.4329 <sup>a</sup>	5.4238 <sup>a</sup>	4.0607 <sup>a,b</sup>	5.2971 <sup>a</sup>
	SD	0.4096	0.184	0.3536	0.2934	0.3054	0.3471
	Region A ( <i>n</i> = 42)						
Mean	3.7527 <sup>1</sup>	3.8392 <sup>1</sup>	3.0905 <sup>1</sup>	5.4835 <sup>1</sup>	4.11261	4.9533 <sup>1</sup>	
SD	0.6361	0.467	0.9664	0.3564	0.3558	0.7137	
B	Nova Esperança ( <i>n</i> = 20)						
	Mean	3.8899 <sup>a</sup>	3.6712 <sup>a,b</sup>	3.4773 <sup>a</sup>	5.6537 <sup>a</sup>	4.2744 <sup>a,b</sup>	5.0994 <sup>a</sup>
	SD	0.793	0.5724	1.22	0.454	0.3402	0.8469
	Londrina ( <i>n</i> = 20)						
	Mean	3.9402 <sup>a</sup>	3.3980 <sup>a,b,c</sup>	2.9169 <sup>a</sup>	5.43583 <sup>a</sup>	3.9828 <sup>a,b</sup>	4.9497 <sup>a</sup>
	SD	0.5018	0.6788	0.7313	0.3342	0.3484	0.4137
	Uraí ( <i>n</i> = 11)						
	Mean	4.1034 <sup>a</sup>	2.7858 <sup>c</sup>	2.7312 <sup>a</sup>	5.3946 <sup>a</sup>	4.0209 <sup>a,b</sup>	5.5468 <sup>a</sup>
	SD	0.8494	0.5455	1.1206	0.4545	0.5435	0.6651
	Santa Mariana ( <i>n</i> = 12)						
Mean	4.2236 <sup>a</sup>	3.2597 <sup>a,b,c</sup>	3.0346 <sup>a</sup>	5.4645 <sup>a</sup>	4.2176 <sup>a,b</sup>	5.5395 <sup>a</sup>	
SD	0.5343	0.6095	1.2284	0.6748	0.5483	0.5267	
Region B ( <i>n</i> = 69)							
Mean	3.9779 <sup>1</sup>	3.3303 <sup>2</sup>	3.0388 <sup>1</sup>	5.4938 <sup>1</sup>	4.1227 <sup>1</sup>	5.2005 <sup>1</sup>	
SD	0.6704	0.7093	1.104	0.4729	0.4408	0.6702	
C	Guapirama ( <i>n</i> = 10)						
	Mean	3.7633 <sup>a</sup>	3.5042 <sup>a,b,c</sup>	3.3478 <sup>a</sup>	5.1683 <sup>a</sup>	4.1595 <sup>a,b</sup>	4.9174 <sup>a</sup>
	SD	0.2922	0.8056	0.8175	0.4919	0.2619	0.4652
	Joaquim Távora ( <i>n</i> = 9)						
	Mean	3.9049 <sup>a</sup>	3.8129 <sup>a,b</sup>	3.5019 <sup>a</sup>	5.4093 <sup>a</sup>	4.5939 <sup>a</sup>	5.223 <sup>a</sup>
	SD	0.649	1.0521	1.3434	0.2304	0.4266	0.8684
	Jacarezinho ( <i>n</i> = 18)						
	Mean	4.1836 <sup>a</sup>	3.50940 <sup>a,b,c</sup>	3.3141 <sup>a</sup>	5.3316 <sup>a</sup>	4.1890 <sup>a,b</sup>	5.0667 <sup>a</sup>
	SD	0.7048	0.4151	1.0337	0.3206	0.2589	0.4701
	Abatiá ( <i>n</i> = 11)						
Mean	3.4802 <sup>a</sup>	3.6362 <sup>a,b,c</sup>	3.2945 <sup>a</sup>	5.5152 <sup>a</sup>	3.8912 <sup>b</sup>	4.9871 <sup>a</sup>	
SD	0.445	0.4175	0.9896	0.5192	0.3196	0.3971	
Region C ( <i>n</i> = 61)							
Mean	3.9251 <sup>1</sup>	3.5687 <sup>1,2</sup>	3.3027 <sup>1</sup>	5.3726 <sup>1</sup>	4.2106 <sup>1</sup>	5.0592 <sup>1</sup>	
SD	0.6343	0.6591	0.9802	0.3991	0.3818	0.5523	

For cities, means in the same column followed by same letter are not statistically different at the level of 5%

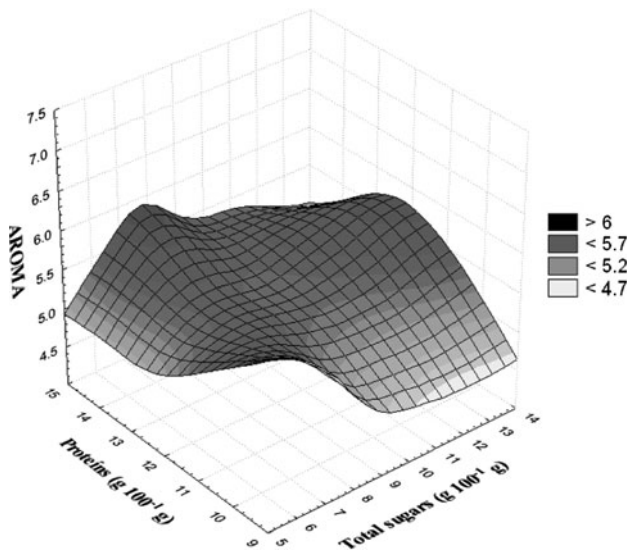
For regions, means in the same column followed by same number are not statistically different at the level of 5%

## Conclusions

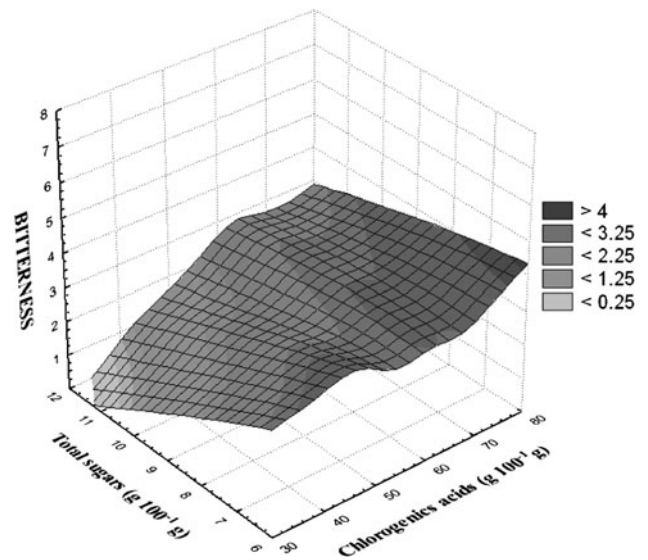
The multilayer perceptron type artificial neural networks were shown to be useful tools to identify coffee samples

because they identified 100% of the samples at the test and validation stages for the division by planting regions. For each databank containing information per city, using the automatic model, the artificial neural network used

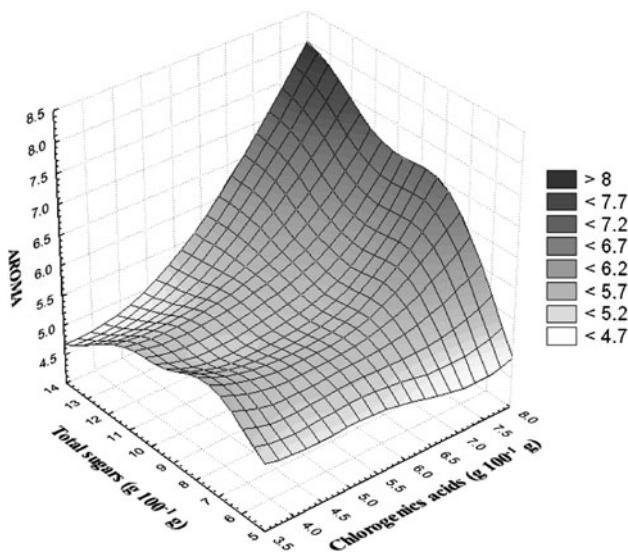




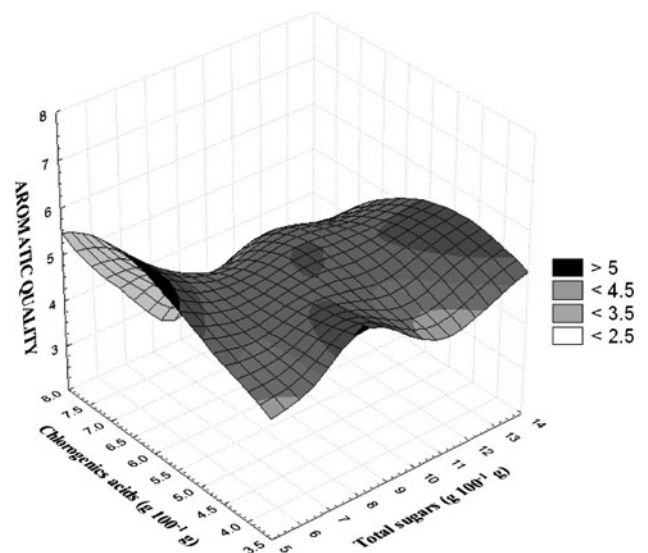
**Fig. 4** Response surface showing the influence of the protein and total sugar contents on coffee aroma regardless of the regions studied



**Fig. 6** Response surface showing the influence of the total sugars and chlorogenic acids on the bitterness of coffee regardless of the regions studied



**Fig. 5** Response surface showing the influence of the total sugar and chlorogenic acid contents on coffee aroma regardless of the regions studied



**Fig. 7** Response surface showing the influence of the total sugar and chlorogenic acid content on the aromatic quality of the coffee regardless of the regions studied

presented 99% accuracy in training and 100% accuracy in the test and validation stages.

Using artificial neural networks, we can efficiently classify coffees from regions and cities in Parana and with this, identify the geographic origin of the coffee and associate quality with the producing location. Furthermore, we can assess the importance of some substances present in coffee in the formation of the sensorial attributes of the beverage.

**Acknowledgments** The authors are grateful to Fundação Araucária, UEL and IAPAR for their financial support.

**References**

1. Bishop CM (1995) Neural networks for pattern recognition. University Oxford, Oxford
2. Haykin S (2001) Redes neurais: princípios e práticas. Bookman, Porto Alegre
3. Hair JF Jr, Anderson RE, Tatham RL, Black WC (2005) Análise fatorial. In: Análise Multivariada de Dados, 5ª ed. Bookman, Porto Alegre

4. Melssen W, Wehrens R, Buydens L (2006) Supervised Kohonen networks for classification problems. *Chemom Intell Lab Syst* 83:99–113
5. Kovács ZL (1996) *Redes neurais Artificiais: Fundamentos e Aplicações*. Acadêmica, São Paulo
6. Mukesh D (1996) Applications of neural computing for process chemists: I. Introduction to neural network. *J Chem Educ* 73:431–433
7. Linder R, Pöpl SJ (2003) A new neural network approach classifies olfactory signals with high accuracy. *Food Qual Preference* 14:435–440
8. Deisingh AK, Stone DC, Thompson M (2004) Applications of electronic noses and tongues in food analysis. *Int J Food Sci Technol* 39:587–604
9. Liao S (2005) Expert system methodologies and applications—a decade review from 1995 to 2004. *Expert Systems Appl* 28:93–109
10. Hammond MH, Riedel CJ, Rose-Pehrsson SL, Williams FW (2004) Training set optimization methods for a probabilistic neural network. *Chem Intell Lab Syst* 71:73–78
11. Braga AP, Carvalho ACPLF, Ludermir TB (2000) *Redes Neurais Artificiais: Teoria e Aplicações*, 1st edn. LTC, Rio de Janeiro
12. Azevedo FM, Brasil LM, Oliveira RCL (2000) *As Redes Neurais com Aplicações em Controle e em Sistemas Especialistas*, 1st edn. Bookstores, Florianópolis
13. Hunter A, Kenedy L, Henry J, Fergusom I (2000) Application of neural networks and sensitivity analysis to improved prediction of trauma survival. *Methods Programs Biomed* 62:11–19
14. Serra F, Guillou CG, Reniero F, Ballarin L, Cantagallo MI, Wieser M, Iyer SS, Héberger K, Vanhaecke F (2005) Determination of geographic origin of green coffee by component analysis of carbon, nitrogen and boron stable isotope ratios. *Rapid Commun Mass Spectrom* 19:2011–2015
15. Anderson K, Smith BW (2002) Chemical profiling to differentiate geographic growing origins of coffee. *J Agric Food Chem* 50:2068–2075
16. Bodt E, Cottrell M, Verleysen M (2002) Statistical tools to assess the reliability of self-organizing maps. *Neural Netw* 15:967–978
17. Looney CG (1996) Advances in feedforward neural networks: demystifying knowledge acquiring black boxes. *IEEE Trans Knowl Data Eng* 8:211–226
18. Ludermir TB, Yamazaki A, Zanchetin C (2006) An optimization methodology for neural networks weights and architectures. *IEEE Trans Knowl Data Eng* 17:1452–1459
19. Windeatt T (2006) Accuracy/diversity and ensemble MLP classifier design. *IEEE Trans Neural Netw* 17:1194–1211
20. StatSoft, Inc (2009) STATISTICA (data analysis software system), version 9.0. <http://www.statsoft.com>
21. Maeztu L, Sanz C, Andueza S, De Penã MP, Bello J (2001) Characterization of espresso coffee aroma by static headspace gc-ms and sensory flavor profile. *J Agric Food Chem* 49:5437–5444
22. De Maria CAB, Moreira RFA, Trugo LC (1999) Componentes voláteis do café torrado. Parte I: compostos heterocíclicos. *Quim Nova* 22:209–217
23. Grosch W (2001) Evaluation of the key odorants of foods by dilute experiments, aroma models and omission. *Chem Senses* 26:533–545
24. Mazzafera P (1999) Chemical composition of defective coffee beans. *Food Chem* 64:547–554
25. Clifford MN (1985) Chemical and physical aspects of green coffee and coffee products. In: Clifford MN, Wilson KC (eds) *Coffee, botany, biochemistry and production of beans and beverage*. Croom Helm, London, pp 305–359
26. Carvalho VD, Chagas SJR, Chalfoun SM, Botrel N, Juste ESG (1994) Relação entre a composição físico-química e química do grão beneficiado e a qualidade de bebida do café. *Pesq Agropec Bras* 29:449–454
27. BRASIL. Ministério da Agricultura Pecuária e Abastecimento. Instrução Normativa n° 8 de 11 June 2003
28. Instituto Adolfo Lutz. Métodos físicos químicos para análise de alimentos. Edição IV, 1ª edição digital, São Paulo, 2008
29. Clifford MN, Wight JC (1976) Chlorogenic acids and other cinnamates nature, occurrence and dietary burden. *J Sci Food Agric* 27:73–84
30. AOAC—Association of Official Agricultural Chemists (1990) *Official methods of analysis*, 15th edn. Washington
31. Southgate DAT (1976) *Determination of food carbohydrates*. Applied Science Publishers, Londres
32. Pavan MA, Bloch MDM, Zempulsk HC, Miyazawa M, Zocoler DC (1992) *Manual de análises químicas de solo e controle da qualidade*. IAPAR, Londrina, Circular Técnica, n. 76
33. Basheer IA, Hajmeer M (2000) Artificial neural networks: fundamentals, computing, design, and application. *J Microbiol Methods* 43:3–31
34. Broyden CG (1970) The convergence of a class of double rank minimization algorithms: 2. The new algorithm. *J Appl Math* 6:222–231
35. Fletcher R (1970) A new approach to variable metric algorithms. *Comput J* 13:317–322
36. Goldfarb D (1970) A family of variable metric methods derived by variational means. *J Comput Math* 24:23–26
37. Shanno DF (1970) Conditioning of quasi-Newton methods for function minimization. *Math Comput* 24:647–650
38. Dai YH (2002) Convergence properties of the BFGS algorithm. *SIAM J Optim* 13:693–701
39. Fernandes SM, Pinto NAVD, Thé PMP, Pereira RGFA, Carvalho VD (2001) Teores de polifenóis, ácido clorogênico, cafeína e proteína em café torrado. *Rev Bras Agrociência* 7:197–199
40. Sarrazin C, Lequère JL, Gretsche C, Liardon R (2000) Representativeness of coffee aroma extracts: a comparison of different extraction methods. *Food Chem* 70:99–106
41. Illy A, Viani R (1996) *Espresso coffee: The chemistry of quality*, 2nd edn. Academic press, San Diego
42. Sivetz M, Desrosier NW (1979) *Physical and chemical aspects of coffee*. Coffee Technology, Westport
43. Ginz M, Balzer HH, Bradbury AGW, Maier HG (2000) Formation of aliphatic acids by carbohydrate degradation during roasting of coffee. *Eur Food Res Tech* 211:404–410
44. Rogers WJ, Michaux S, Bastin M, Bucheli P (1999) Changes to the content of sugar, sugar alcohols, myo-inositol, carboxylic acid and inorganic anions in developing grains from different varieties of Robusta (*Coffea canephora*) and arabica (*Coffea arabica*) coffees. *Plant Sci* 149:115–123
45. Vasconcelos AL, Franca AS, Gloria MBA, Mendonça JCF (2007) A comparative study of chemical attributes and levels of amines in defective green and roasted coffee beans. *Food Chem* 100:26–32
46. Farah A, Monteiro MC, Calado V, Franca AS, Trugo LC (2006) Correlation between cup quality and chemical attributes of Brazilian coffee. *Food Chem* 98:373–380
47. Moon JK, Yoo HS, Shibamoto T (2009) Role of roasting conditions in the level of chlorogenic acid content in coffee beans: correlation with coffee acidity. *J Agric Food Chem* 57:5365–5369
48. Jiang D, Peterson DG (2010) Role of hydroxycinnamic acids in food flavor: a brief overview. *Phytochem Rev* 9:187–193

49. Frank O, Blumberg S, Kunert C, Zehentbauer G, Hofmann T (2007) Structure determination and sensory analysis of bitter-tasting 4-vinylcatechol oligomers and their identification in roasted coffee by means of LC-MS/MS. *J Agric Food Chem* 55:1945–1954
50. Montavon P, Duruz E, Rumo G, Pratz G (2003) Evolution of green coffee protein profiles with maturation and relationship to coffee cup quality. *J Agric Food Chem* 51:2328–2334